BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号

特開平8－263224

(43)公開日　平成８年(1996)10月11日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ　3/06	５４０		Ｇ０６Ｆ　3/06	５４０	

審査請求　未請求　請求項の数41　ＯＬ　（全 20 頁）

(21)出願番号　特願平8－25038

(22)出願日　平成８年(1996)２月13日

(31)優先権主張番号　３９６０４６

(32)優先日　1995年２月28日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　390009531
インターナショナル・ビジネス・マシーンズ・コーポレイション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国10504、ニューヨーク州　アーモンク　（番地なし）

(72)発明者　ポール・ホッジス
アメリカ合衆国95120、カリフォルニア州　サン・ホセ、モンテベルデ・ドライブ　6154

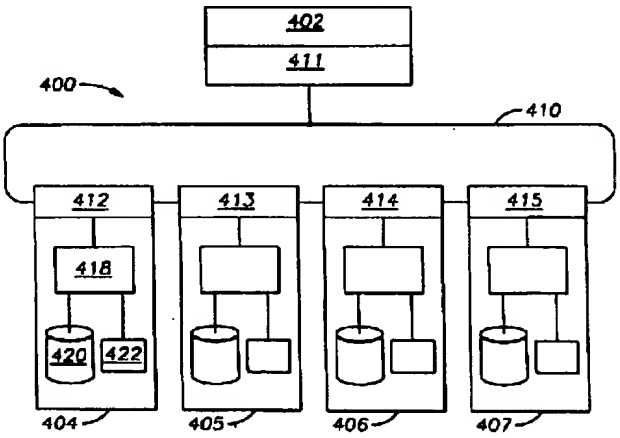
(74)代理人　弁理士　合田　潔　（外２名）

(54)【発明の名称】　局所ＸＯＲ機能を有するデータ記憶システム

(57)【要約】

【課題】　複数のＸＯＲ演算を記憶装置間で分散して実行するデータ記憶システムを提供する。

【解決手段】　データ記憶システムが制御装置及び複数のＤＡＳＤのストリングを含み、各ＤＡＳＤが制御装置からのコマンドに応答して、ＸＯＲ機能及び他のＤＡＳＤへのデータ転送などの局所オペレーションを実行する。直列または並列バスを含むバスが、制御装置及びＤＡＳＤを電気的に相互接続する。各ＤＡＳＤは記憶装置、インタフェース、バッファ、及びプロセッサを含む。バッファはインタフェース及び記憶装置から選択的にデータを受信する。他のオペレーションに加え、プロセッサは制御装置からバッファ・インタフェースへのコマンドに応答して、バッファ内の選択データ項目に対してＸＯＲ演算を実行する。プロセッサは更に、演算結果を別のＤＡＳＤまたは制御装置に転送しうる。



1

【特許請求の範囲】

【請求項１】マシン読出し可能データを記憶する装置で使用されるＤＡＳＤであって、前記装置が制御装置、複数のＤＡＳＤ、及び前記制御装置と前記ＤＡＳＤを電気的に相互接続するバスを含み、前記複数のＤＡＳＤのうちの１つが、前記制御装置の各オペレーションにおけるパリティＤＡＳＤとして指定されるものにおいて、

マシン読出し可能データを記憶する記憶装置と、

前記バスに電気的に接続されるインタフェースと、

前記インタフェース及び前記記憶装置に動作的に接続される少なくとも１つのバッファと、

前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介してコマンドを受信し、該コマンドに応答して、前記バッファ内に記憶されるデータ項目を選択し、前記選択データ項目に対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を前記バスを介して、前記コマンドにより指定される、前記ＤＡＳＤの１つまたは前記制御装置を含む宛先に送信するプロセッサと、

を含む、ＤＡＳＤ。

【請求項２】前記プロセッサがマイクロプロセッサを含む、請求項１記載のＤＡＳＤ。

【請求項３】前記プロセッサがＡＳＩＣを含む、請求項１記載のＤＡＳＤ。

【請求項４】前記プロセッサが個別回路要素のアセンブリから成る論理回路を含む、請求項１記載のＤＡＳＤ。

【請求項５】前記記憶装置が、少なくとも１つの磁気記録ディスクを含むディスク・ドライブを含む、請求項１記載のＤＡＳＤ。

【請求項６】前記インタフェースが直列インタフェースを含む、請求項１記載のＤＡＳＤ。

【請求項７】前記インタフェースが並列インタフェースを含む、請求項１記載のＤＡＳＤ。

【請求項８】バスにより相互接続される複数のメンバを含むデータ記憶システムにおいて使用されるＤＡＳＤであって、前記メンバが前記制御装置及び複数のＤＡＳＤを含むものにおいて、

前記バスに電気的に接続されるインタフェースと、

データ記憶装置と、

前記インタフェース及び前記記憶装置に動作的に接続され、それらから受信されるデータを選択的に記憶するバッファと、

前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介して受信されるコマンドに応答して、次のステップ、すなわち、前記バッファと前記インタフェース間でデータを選択的に転送するステップ、前記記憶装置と前記バッファ間でデータを交換するステップ、前記バッファ内に含まれる前記選択データ項目に対してＸＯＲ演算を実行するステップ、及び前記インタフェースから選択される前記メンバにデータを転送するステップの少なくとも１つを実行するようにプログラムされるプロセッサと、

を含む、ＤＡＳＤ。

2

【請求項９】前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記記憶装置から選択データを読出し、該選択データを選択される前記メンバに転送するステップを実行するようにプログラムされる、請求項８記載のＤＡＳＤ。

【請求項１０】前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記インタフェースにより受信される前記選択データを前記記憶装置に書込むステップを実行するようにプログラムされる、請求項８記載のＤＡＳＤ。

【請求項１１】前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記記憶装置から選択データを読出し、前記インタフェースにより受信されるデータ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送するステップを実行するようにプログラムされる、請求項８記載のＤＡＳＤ。

【請求項１２】前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記記憶装置から前記選択データを読出し、前記インタフェースにより受信されるデータ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送し、前記インタフェースにより受信される前記データを前記記憶装置に書込むステップを実行するようにプログラムされる、請求項８記載のＤＡＳＤ。

【請求項１３】前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記記憶装置から前記選択データを読出し、前記インタフェースにより受信されるデータ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を前記記憶装置に書込むステップを実行するようにプログラムされる、請求項８記載のＤＡＳＤ。

【請求項１４】前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記インタフェースにより受信される前記選択データを前記記憶装置に書込み、前記選択データを選択される前記メンバに転送するステップを実行するようにプログラムされる、請求項８記載のＤＡＳＤ。

【請求項１５】前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記インタフェースにより受信される２つのデータ項目に対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送し、選択される一方の前記受信データ項目を前記記憶装置に書込むステップを実行するようにプログラムされる、請求項８記載のＤＡＳＤ。

【請求項１６】マシン読出し可能データを記憶するシステムであって、

制御装置及び複数のＤＡＳＤを含む複数のメンバと、

前記メンバを電気的に相互接続してループを形成するバスと、
を含み、前記の各ＤＡＳＤが、
マシン読出し可能データを記憶する記憶装置と、
前記バスに電気的に接続されるインタフェースと、
前記インタフェース及び前記記憶装置に動作的に接続され、前記インタフェース及び前記記憶装置から受信されるデータを選択的に記憶する少なくとも１つのバッファと、
前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介してコマンドを受信し、所定のコマンドに応答して、前記バッファ内に記憶される前記データ項目を選択し、前記選択データ項目に対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を前記バスを介して、選択される前記メンバを含む宛先に送信するＸＯＲ発生器と、
を含む、システム。
【請求項１７】前記プロセッサがマイクロプロセッサを含む、請求項１６記載のシステム。
【請求項１８】前記プロセッサがＡＳＩＣを含む、請求項１６記載のシステム。
【請求項１９】前記プロセッサが個別回路要素のアセンブリから成る論理回路を含む、請求項１６記載のシステム。
【請求項２０】前記記憶装置が、少なくとも１つの磁気記録ディスクを含むディスク・ドライブを含む、請求項１６記載のシステム。
【請求項２１】前記インタフェースが直列インタフェースを含む、請求項１６記載のシステム。
【請求項２２】前記インタフェースが並列インタフェースを含む、請求項１６記載のシステム。
【請求項２３】前記複数のＤＡＳＤがＲＡＩＤプロトコルに従い構成される、請求項１６記載のシステム。
【請求項２４】前記ＲＡＩＤプロトコルがＲＡＩＤ－５を含む、請求項２３記載のシステム。
【請求項２５】複数のＸＯＲ演算の実行を要求するタスクを実行するように構成されるデータ記憶システムであって、
コマンドを転送する制御装置と、
前記コマンドに応答して、前記ＸＯＲ演算の実行をＤＡＳＤ間で分散することにより、前記タスクを実行するＤＡＳＤアレイと、
前記ＤＡＳＤ及び前記制御装置を相互接続するバスと、
を含む、データ記憶システム。
【請求項２６】データ記憶システムを動作させる方法であって、
制御装置及び複数のＤＡＳＤのストリングを含む複数のメンバと、
前記複数のメンバを電気的に相互接続するバスと、
を含み、各ＤＡＳＤが、

前記バスに電気的に接続されるインタフェースと、
データ記憶装置と、
前記インタフェース及び前記記憶装置に動作的に接続されるバッファと、
前記バッファに動作的に接続されるプロセッサと、
を含み、
使用可能な前記ＤＡＳＤを識別するステップと、
前記制御装置から前記識別されたＤＡＳＤに個々にコマンドを送信するステップと、
前記コマンドに応答して、データが前記識別されたＤＡＳＤを所定順序でデイジー・チェイン転送されるように、前記識別されたＤＡＳＤを動作させるステップと、
を含み、データがイニシエータＤＡＳＤから始まり、複数の中間ＤＡＳＤを経由してレシーバＤＡＳＤに至る前記各ＤＡＳＤにより、順次転送及び受信され、前記の各中間ＤＡＳＤが、前記デイジー・チェインに沿って受信される前記データを前記所定順序に従う次のＤＡＳＤに転送する以前に、前記受信データ項目及び別のデータ項目に対して、ＸＯＲ演算を実行する、方法。
【請求項２７】前記別のデータ項目が前記制御装置から受信される前記データ項目を含む、請求項２６記載の方法。
【請求項２８】前記別のデータ項目が前記中間ＤＡＳＤのそれぞれの前記記憶装置から読出される前記データ項目を含む、請求項２６記載の方法。
【請求項２９】前記制御装置からの少なくとも１つの前記所定コマンドに応答して、選択ＤＡＳＤの前記記憶装置から前記データを読出し、読出したデータを選択される前記メンバに転送するように、前記選択ＤＡＳＤをオペレートするステップを含む、請求項２６記載の方法。
【請求項３０】前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記インタフェースにより受信される前記データを前記選択ＤＡＳＤの前記記憶装置に書込むように、前記選択ＤＡＳＤをオペレートするステップを含む、請求項２６記載の方法。
【請求項３１】前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記記憶装置から前記選択データを読出し、前記選択ＤＡＳＤの前記インタフェースにより受信される前記データ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送するように、前記選択ＤＡＳＤをオペレートするステップを含む、請求項２６記載の方法。
【請求項３２】前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記記憶装置から前記選択データを読出し、前記選択ＤＡＳＤの前記インタフェースにより受信される前記データ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送し、前記選択ＤＡＳＤの前記インタフェースにより受信された前記デー

タを該選択ＤＡＳＤの前記記憶装置に書込むように、前記選択ＤＡＳＤをオペレートするステップを含む、請求項２６記載の方法。

【請求項３３】前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記記憶装置から前記選択データを読出し、前記選択ＤＡＳＤの前記インタフェースにより受信される前記データ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を前記選択ＤＡＳＤの前記記憶装置に書込むように、前記選択ＤＡＳＤをオペレートするステップを含む、請求項２６記載の方法。

【請求項３４】前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記インタフェースにより受信される前記選択データを前記選択ＤＡＳＤの前記記憶装置に書込み、前記受信データを選択される前記メンバに転送するように、前記選択ＤＡＳＤをオペレートするステップを含む、請求項２６記載の方法。

【請求項３５】前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記インタフェースにより受信される２つの前記データ項目に対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送し、選択される一方の前記受信データ項目を前記選択ＤＡＳＤの前記記憶装置に書込むように、前記選択ＤＡＳＤをオペレートするステップを含む、請求項２６記載の方法。

【請求項３６】制御装置、ＤＡＳＤアレイ、及び前記ＤＡＳＤ及び前記制御装置を相互接続するバスを含むデータ記憶システムにおいて、複数のＸＯＲ演算の実行を要求するタスクを実行するように、前記データ記憶システムを動作させる方法であって、

前記制御装置から前記バスを介して前記ＤＡＳＤにコマンドを転送するステップと、

前記コマンドに応答して、前記ＸＯＲ演算の実行を前記ＤＡＳＤ間で分散することにより、前記タスクを実行するように、前記ＤＡＳＤをオペレートするステップと、

を含む、方法。

【請求項３７】前記オペレーティング・ステップが、データを前記ＤＡＳＤ間で選択順に順次転送するように、前記ＤＡＳＤをオペレートするステップを含み、前記複数のＤＡＳＤが前記選択順序に従い前記データを別のＤＡＳＤに転送する以前に、該データに対して前記ＸＯＲ演算を実行する、請求項３６記載の方法。

【請求項３８】前記オペレーティング・ステップが再生オペレーションを含み、該再生オペレーションが、

第１のＤＡＳＤの第１の記憶ロケーションから第１のデータ・ブロックを読出し、前記第１のデータ・ブロックを第２のＤＡＳＤに転送するように、前記第１のＤＡＳＤをオペレートするステップと、

前記第２のＤＡＳＤをオペレートするステップであって、

前記第１のデータ・ブロックを前記第１のＤＡＳＤから受信するステップと、

前記第２のＤＡＳＤの記憶ロケーションから、前記第１のデータ・ブロックに対応する第２のデータ・ブロックを読出すステップと、

前記第１及び前記第２のデータ・ブロックに対してＸＯＲ演算を実行し、第１のＸＯＲ結果を生成するステップと、

前記第１のＸＯＲ結果を第３のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、

前記第３のＤＡＳＤをオペレートするステップであって、

前記第１のＸＯＲ結果を前記第２のＤＡＳＤから受信するステップと、

前記第３のＤＡＳＤの記憶ロケーションから、前記第１及び前記第２のデータ・ブロックに対応する第３のデータ・ブロックを読出すステップと、

前記第１のＸＯＲ結果及び前記第３のデータ・ブロックに対してＸＯＲ演算を実行し、第２のＸＯＲ結果を生成するステップと、

前記第２のＸＯＲ結果を前記制御装置に転送するステップと、を含む前記オペレーティング・ステップと、

を含む、請求項３６記載の方法。

【請求項３９】前記オペレーティング・ステップが復元オペレーションを含み、該復元オペレーションが、

前記第１のＤＡＳＤの前記第１の記憶ロケーションから第１のデータ・ブロックを読出すように、前記第１のＤＡＳＤをオペレートするステップと、

第２のＤＡＳＤをオペレートするステップであって、

前記第１のデータ・ブロックを前記第１のＤＡＳＤから受信するステップと、

前記第２のＤＡＳＤの記憶ロケーションから、前記第１のデータ・ブロックに対応する前記第２のデータ・ブロックを読出すステップと、

前記第１及び前記第２のデータ・ブロックに対してＸＯＲ演算を実行し、第１のＸＯＲ結果を生成するステップと、

前記第１のＸＯＲ結果を前記第３のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、

前記第３のＤＡＳＤをオペレートするステップであって、

前記第１のＸＯＲ結果を前記第２のＤＡＳＤから受信するステップと、

前記第３のＤＡＳＤの記憶ロケーションから、前記第１及び前記第２のデータ・ブロックに対応する前記第３のデータ・ブロックを読出すステップと、

前記第１のＸＯＲ結果及び前記第３のデータ・ブロックに対してＸＯＲ演算を実行し、第２のＸＯＲ結果を生成

するステップと、
前記第２のＸＯＲ結果を第４のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、
前記第４のＤＡＳＤをオペレートするステップであって、
前記第２のＸＯＲ結果を前記第３のＤＡＳＤから受信するステップと、
前記第２のＸＯＲ結果を前記第４のＤＡＳＤの記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、
を含む、請求項３６記載の方法。
【請求項４０】前記オペレーティング・ステップが更新オペレーションを含み、該更新オペレーションが、
前記第１のデータ・ブロックを前記制御装置から前記第１のＤＡＳＤの前記第１の記憶ロケーションに転送するステップと、
前記第１のＤＡＳＤをオペレートするステップであって、
前記第１のデータ・ブロックを前記制御装置から受信するステップと、
前記第１のＤＡＳＤの前記記憶ロケーションから、前記第１のデータ・ブロックに対応する第２のデータ・ブロックを読出すステップと、
前記第１及び前記第２のデータ・ブロックに対してＸＯＲ演算を実行し、第１のＸＯＲ結果を生成するステップと、
前記第１のＸＯＲ結果を前記第２のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、
前記第２のＤＡＳＤをオペレートするステップであって、
前記第１のＸＯＲ結果を前記第１のＤＡＳＤから受信するステップと、
前記第２のＤＡＳＤの前記記憶ロケーションから、前記第１及び前記第２のデータ・ブロックに対応する前記第３のデータ・ブロックを読出すステップと、
前記第２及び前記第３のデータ・ブロックに対してＸＯＲ演算を実行し、第２のＸＯＲ結果を生成するステップと、
前記第２のＸＯＲ結果を前記第３のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、
前記第３のＤＡＳＤをオペレートするステップであって、
前記第２のＸＯＲ結果を前記第２のＤＡＳＤから受信するステップと、
前記第２のＸＯＲ結果を前記第３のＤＡＳＤの記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、
を含む、請求項３６記載の方法。

【請求項４１】前記オペレーティング・ステップがストライプ書込みオペレーションを含み、該ストライプ書込みオペレーションが、
第１のデータ・ブロックを前記制御装置から前記第１のＤＡＳＤの前記第１の記憶ロケーションに転送するステップと、
前記第１のＤＡＳＤをオペレートするステップであって、
前記第１のデータ・ブロックを前記制御装置から受信するステップと、
前記第１のデータ・ブロックを前記第１のＤＡＳＤの記憶ロケーションに書込むステップと、
前記第１のデータ・ブロックを前記第２のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、
前記第２のＤＡＳＤをオペレートするステップであって、
前記第２のデータ・ブロックを前記制御装置から受信するステップと、
前記第２のデータ・ブロックを前記第２のＤＡＳＤの記憶ロケーションに書込むステップと、
前記第１及び前記第２のデータ・ブロックに対してＸＯＲ演算を実行し、第１のＸＯＲ結果を生成するステップと、
前記第１のＸＯＲ結果を前記第３のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、
前記第３のＤＡＳＤをオペレートするステップであって、
前記第３のデータ・ブロックを前記制御装置から受信するステップと、
前記第３のデータ・ブロックを前記第３のＤＡＳＤの記憶ロケーションに書込むステップと、
前記第１のＸＯＲ結果を前記第２のＤＡＳＤから受信するステップと、
前記第１のＸＯＲ結果及び前記第３のデータ・ブロックに対してＸＯＲ演算を実行し、第２のＸＯＲ結果を生成するステップと、
前記第２のＸＯＲ結果を前記第４のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、
前記第４のＤＡＳＤをオペレートするステップであって、
前記第２のＸＯＲ結果を前記第３のＤＡＳＤから受信するステップと、
前記第２のＸＯＲ結果を前記第４のＤＡＳＤの記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、
を含む、請求項３６記載の方法。

【発明の詳細な説明】

【０００１】
【発明の属する技術分野】本発明はＲＡＩＤ（redundant arrays of inexpensive disks）などのデータ記憶媒体のアレイに関し、特に、制御装置と、各々が受信データを処理し、処理データを異なる記憶装置または制御装置に送信するための局所排他的論理和（ＸＯＲ）発生器を含む複数のデータ記憶装置とを使用するデータ記憶システムに関する。従って、本発明のデータ記憶システムは、複数のＸＯＲ演算の実行を要求するタスクを、それらのＸＯＲ演算を記憶装置間で分散して実行するように動作する。
【０００２】
【従来の技術】多くの技術者が、ＲＡＩＤなどの拡張データ記憶システムを使用することにより、コンピュータの記憶能力を改善しようと努めてきた。一般的に、ＲＡＩＤ技術は、複数の同一の直接アクセス記憶装置（ＤＡＳＤ）を使用することにより、デジタル・データを記憶する。ＲＡＩＤ技術の概念は既知であり、ＲＡＩＤ装置の多くの異なる変形が開発されてきた。例えば、"A Case for Redundant Arraysof Inexpensive Disks(RAID)"、Report No.UCB/CSD 87/391、December 1987（Patersonらによる"A Case for Redundant Arrays of Inexpensive Disks(RAID)"、Proceedings of the 1988 ACM SIGMOD Conference on Management of Data, Chicago, IL, June 1988としても引用される）を参照されたい。
【０００３】典型的ＲＡＩＤ装置は、ＲＡＩＤ制御装置と複数のＤＡＳＤとを含む。各ＤＡＳＤはマイクロプロセッサ及び様々なメモリ・バッファなどの特定の電子回路とと共に、複数のデータ記憶ディスクを含む。各ＤＡＳＤの特定部分が、他のＤＡＳＤに対応するパリティ・ビットを記憶するために割当てられる。しかしながら、説明を容易にするために、ＲＡＩＤはしばしば、パリティ情報の記憶を専用とする特定のＤＡＳＤを有するように述べられる。
【０００４】ＲＡＩＤシステムでは、データは通常、アレイの異なるＤＡＳＤの対応ロケーションに渡り、データをストライピングすることにより記憶される。記憶データのパリティ・ビットは、データにＸＯＲ機能を適用することにより生成される。従って、あるＤＡＳＤが故障すると、喪失データが、残りのＤＡＳＤからのデータ及び対応パリティ・データにＸＯＲ機能を適用することにより復元される。
【０００５】多くのＲＡＩＤ記憶システムでは、次のような多数の標準機能が存在する。
１．順次書込み
２．更新書込み
３．キャッシュ・データによる更新
４．故障データＤＡＳＤ状況の更新
５．故障パリティＤＡＳＤ状況の更新
６．喪失データの再生（regenerate）
７．喪失データの復元（rebuild）
８．データの読出し
【０００６】順次書込みは、単に、データ及びその対応パリティ・ビットをＤＡＳＤに記憶する。特に、書込みでは、ＤＡＳＤの対応ロケーションに渡り、データの複数ブロック・グループをストライピングし、ストライプ・データに対応するパリティ・ビットのブロックを計算し、パリティ・ブロックをパリティＤＡＳＤに記憶する。パリティ・ブロックは、例えば書込みデータのブロックに対してＸＯＲ演算を実行することにより、計算される。
【０００７】更新書込みは、新たなデータを"ターゲット"ＤＡＳＤに書込み、その新たなデータに対応するパリティ・ビットを更新する。特に更新では、新たなデータ・ブロックを制御装置からターゲットＤＡＳＤに書込み、更新データを反映する更新パリティ・ブロックを計算し、更新パリティ・ブロックを記憶する。更新パリティ・ブロックは、例えば、新たな（更新）データ・ブロック、更新データ・ブロックに対応する旧データ・ブロック、及び対応する旧パリティ・ブロックにＸＯＲ演算を実行することにより計算される。
【０００８】キャッシュ・データによる更新は、制御装置またはＤＡＳＤの１つのキャッシュ・メモリから、ターゲットＤＡＳＤへ新たなデータを書込み、データ及びそのパリティ・ビットを記憶する。より詳細には、キャッシュ・データによる更新は、キャッシュ・データ・ブロックのターゲットＤＡＳＤへの書込み、更新データを反映する更新パリティ・ブロックの計算、及び更新パリティ・ブロックの記憶を含む。更新パリティ・ビットは、例えば、新たな（更新）キャッシュ・データ、旧キャッシュ・データ、及びキャッシュ・データに対応する旧パリティ・ビットに対して、ＸＯＲ演算を実行することにより計算される。
【０００９】故障データＤＡＳＤ状況の更新は、新たなデータによりあるＤＡＳＤが更新されるべきであるが、そのＤＡＳＤが故障している状況において発生する。データが故障ＤＡＳＤに書込めないので、ＲＡＩＤ装置がせいぜい実行できることは、本来書込まれるべきデータを反映するように、パリティＤＡＳＤを更新することである。従って、このプロセスは、非故障データＤＡＳＤから"旧"データ・ブロックを読出し、このデータ及び新データにもとづきパリティ・ブロックを計算し、更新パリティ・ビットをパリティＤＡＳＤ上に記憶する。パリティ・ビットはもちろん、旧データ及び新データにＸＯＲ演算を実行することにより計算される。
【００１０】故障パリティＤＡＳＤ状況の更新は、データ・ブロックがターゲットＤＡＳＤにおいて更新されるが、パリティＤＡＳＤが故障している状況において発生する。このオペレーションでは、更新データ・ブロックが単にターゲットＤＡＳＤに書込まれる。パリティＤＡ

ＳＤを更新するステップは省略される。

【００１１】故障ＤＡＳＤからの喪失データの再生（regenerating）は、故障ＤＡＳＤからデータを復元し、それをＲＡＩＤ制御装置に提供する。特に、このプロセスは全ての非故障ＤＡＳＤから対応データ・ブロックを読出し、読出されたデータ・ブロックにもとづき喪失データ・ブロックを復元し、復元されたデータ・ブロックを制御装置に転送する。喪失データ・ブロックは、例えば、非故障ＤＡＳＤから読出されるデータ・ブロックに対して、ＸＯＲ演算を実行することにより復元される。

【００１２】復元（rebuilding）は、以前に故障した（現在は使用可能な）ターゲットＤＡＳＤからデータ・ブロックを復元（reconstruct）し、復元されたデータ・ブロックをターゲットＤＡＳＤに書込み、パリティ・ブロックを更新する。

【００１３】データの読出しは、複数のＤＡＳＤの１つに記憶されるデータ・ブロックを読出す。

【００１４】従来、技術者達はＲＡＩＤにおける複数の異なる変形を開発した。これらの多くは、ここでの議論で使用されるＲＡＩＤ－５機構に向けられる。ＲＡＩＤシステムは通常、集中化または分散構成により構成される。集中化構成では、ＲＡＩＤ制御装置がＤＡＳＤのために多くの機能を実行し、上述の様々なＲＡＩＤ機能を実行する支配的な役割を演じる。この場合、制御装置はＲＡＩＤシステムの"頭脳"のごとく作用し、コマンドを個々のＤＡＳＤに提供し、ＤＡＳＤから情報を収集し、ＤＡＳＤ内でのデータの処理及び記憶を制御する。制御装置は例えばパリティ・ビットの生成を管理し、ＤＡＳＤの１つが故障した場合のデータの復元を制御する。集中化ＲＡＩＤシステムの重要な特長の１つは、制御装置がドライブに書込まれるデータまたはそこから取り出されるデータに対して実行されなければならない全てのＸＯＲ機能を実行することである。

【００１５】一方、分散システムは、かなりの責任を個々のＤＡＳＤに帰属させる。分散システムでは、制御装置はある程度ＤＡＳＤを管理するが、通常、ＤＡＳＤの１つが、他のＤＡＳＤの特定のオペレーションを管理する重要な役割を担う。本発明を理解する上で有用な複数の原理を用いる分散ＲＡＩＤシステムの１つは、ここでは"外部ＸＯＲ"アプローチとして参照される。すなわち、ＸＯＲ機能は制御装置にとって"外部的"であるが、ＤＡＳＤにとっては局所的である。こうした局所計算は、喪失データの復元、パリティ計算などを支援するために、順次結合されうる。図１に示されるように、外部ＸＯＲアプローチは、複数のＤＡＳＤ１０４乃至１０７に接続される中央制御装置１０２を用いるＲＡＩＤ装置１００に関わる。各ＤＡＳＤ１０４乃至１０７は、バッファ、ＸＯＲ回路、及びディスク・ドライブなどの記憶装置を含む。例えば、ＤＡＳＤ１０４は、バッファ１０４ａ、ＸＯＲ回路１０４ｂ、及びディスク・ドライブ１０４ｃを含む。

【００１６】図１では、各ＤＡＳＤ１０４乃至１０７が、バス１１０を介して、他のＤＡＳＤ１０４乃至１０７及び制御装置１０２に接続される。外部ＸＯＲ機構の下で、制御装置１０２はイニシエータＤＡＳＤを指定し、単一のＳＣＳＩコマンドを指定イニシエータＤＡＳＤに発行することにより、アレイ・オペレーションを開始する。図１の例では、イニシエータＤＡＳＤはＤＡＳＤ１０７である。イニシエータＤＡＳＤ１０７は、制御装置１０２により要求されるオペレーションを達成するＳＣＳＩイニシエータの役割を担う。すなわち、イニシエータＤＡＳＤ１０７は、制御装置１０２にとって外部的にオペレーションを達成し、終了時に終了ステータスを制御装置１０２に報告する。一般に、エラー回復はＤＡＳＤの役割である。しかしながら、ＤＡＳＤレベルにおいて回復不能なエラーに対しては、エラー回復責任を制御装置１０２に転嫁するプロシージャが存在する。

【００１７】外部ＸＯＲアプローチは、多数の魅力的な特長を有する。例えば、このアプローチは、ＲＡＩＤ制御装置の作業負荷を軽減する。また、ＸＯＲ機能の局所化は、特殊ＲＡＩＤ制御装置の必要を回避しうり、特定のケースでは、ＲＡＩＤ装置が一般のＳＣＳＩ制御装置と互換にさえなりうる。

【００１８】しかしながら、特定のアプリケーションでは、外部ＸＯＲアプローチの下でのシステム性能が、ユーザの期待にそぐわないことがありうる。例えば、制御装置１０２が複数のイニシエータを指定する場合、ＤＡＳＤ１０４乃至１０７が、同時に複数のイニシエータによる一貫性の無い制御に晒されうる。特に、制御装置によるコマンドの発行と、任意のＤＡＳＤによるそのコマンドの受信との間に遅延が存在しうる。従って、連続的なコマンドが予測不能かつ不正な順序で受信され、実行される機会が存在しうる。特定のケースでは、これはデータの保全性を低下しうる。

【００１９】外部ＸＯＲアプローチの別の潜在的欠点は、計算効率に関わる。特にこのアプローチは、特定のＲＡＩＤオペレーションにおいて過剰なタスクを実行する。例えば、図２は順次書込みオペレーションの間の、外部ＸＯＲ装置２００の振舞いを示す。外部ＸＯＲ装置２００は制御装置２０２、データＤＡＳＤ２０４乃至２０６、及びパリティＤＡＳＤ２０８を含む。外部ＸＯＲ装置２００は、順次書込みオペレーションを次のように実行する。第１に、制御装置２０２が、ＸＯＲデータ書込み（ＸＤＷ）コマンド２０４ａ乃至２０６ａを、それぞれＤＡＳＤ２０４乃至２０６に発行する。各書込みコマンドと共に、制御装置２０２は、対応するＤＡＳＤに書込まれるデータ・ブロックを提供する。データ・ブロックは２０４ｂ乃至２０６ｂにより示される。

【００２０】コマンド及びデータ・ブロックの受信に際し、ＤＡＳＤ２０４乃至２０６は、適切なデータをそれ

らのディスク・ドライブに書込む。更に各ＤＡＳＤは、ＸＯＲ機能を新たなデータ及び置換される旧データに適用する。これらのＸＯＲ演算の結果は、後にパリティ・ビットを生成するために使用されるデータを提供し、パリティ・ビットはパリティＤＡＳＤ２０８により記憶される。従って、次のステップでは、ＤＡＳＤ２０４乃至２０６は、パリティＤＡＳＤ２０８に、ＸＯＲパリティ書込み（ＸＰＷ）コマンド２０４ｄ乃至２０６ｄと共に、計算されたパリティ・ビット２０４ｃ乃至２０６ｃを送信する。次に、パリティＤＡＳＤ２０８は受信データから最終パリティ・ビットを計算し、これがパリティＤＡＳＤ２０８のディスク・ドライブに書込まれる。後述の説明で示されるように、外部ＸＯＲアプローチは、図２により示されるような状況において、過剰なオペレーションを要求する。

【００２１】更に、外部ＸＯＲアプローチの別の潜在的欠点は、特定のオペレーションが過度な負荷をイニシエータＤＡＳＤに課すことである。一例として、図３に示される外部ＸＯＲ装置２００により実行される再生オペレーションが挙げられる。再生オペレーションは、故障ＤＡＳＤ３００上に配置されるために取り出せない"ミッシング（missing）"・データ・ブロックを復元するために実行される。再生コマンドはまた再生データを制御装置２０２に提供する。

【００２２】最初に制御装置２０２は、再生コマンド３０２をイニシエータＤＡＳＤに発行する。図３の例では、イニシエータＤＡＳＤはＤＡＳＤ２０８である。イニシエータＤＡＳＤ２０８は次に、ミッシング・データ・ブロックに対応するデータ・ブロックを取り出すために、読出しコマンド３０４乃至３０６を非故障ＤＡＳＤ２０４乃至２０６に送信する。それに応答して、各ＤＡＳＤ２０４乃至２０６は、イニシエータＤＡＳＤ２０８に要求データ・ブロック３０８乃至３１０を送信する。イニシエータＤＡＳＤ２０８はまた、自身のディスク・ドライブからも、ミッシング・データ・ブロックに対応するデータ・ブロックを取り出す。次に制御装置２０２がイニシエータＤＡＳＤ２０８に、ＸＯＲデータ読出し（ＸＤＲ）コマンドを送信する。それに応答して、イニシエータＤＡＳＤ２０８はデータ・ブロックに対してＸＯＲ演算を実行して、ミッシング・データ・ブロックを効果的に再生し、再生データ・ブロックを制御装置２０２に返却する。

【００２３】

【発明が解決しようとする課題】上述のように、外部ＸＯＲアプローチは再生オペレーションの実行において、イニシエータＤＡＳＤ２０８に大きく頼る。イニシエータＤＡＳＤ２０８はコマンドを他の各ＤＡＳＤに送信し、そこからデータ・ブロックを受信し、自身のデータ・ブロックの１つを取り出し、全ての抽出データ・ブロックに対してＸＯＲ演算を実行し、ＸＯＲ結果を制御装置２０２に返却しなければならない。結果的に、イニシエータＤＡＳＤ２０８は、外部ＸＯＲ装置２００のオペレーションを多大に遅延する障害を有しうる。イニシエータＤＡＳＤ２０８の重負荷により、ＸＯＲ演算の実行に相当な計算時間が必要とされる。加えて、この重負荷は、イニシエータＤＡＳＤ２０８がメモリ・バッファ及びバッファ・サイクルなどの、追加のハードウェアを使用することを要求する。

【００２４】更に、外部ＸＯＲ構成は、制御装置からは多大な作業負荷を免除するが、イニシエータＤＡＳＤに課せられる高度な責任が、事実上、イニシエータＤＡＳＤ自体を"制御装置"たらしめる。従って、イニシエータＤＡＳＤが、他のＤＡＳＤと通信し、それらを管理するためのマイクロコード及び制御回路などの、必要な知能を含まねばならない。

【００２５】

【課題を解決するための手段】一般的に、本発明の１つの態様は、マシン読出し可能データを記憶するデータ記憶システムに関する。このシステムは制御装置、複数ＤＡＳＤ、及び制御装置とＤＡＳＤを電気的に相互接続するバスを含む。各ＤＡＳＤはマシン読出し可能データを記憶する記憶装置と、バスに電気的に接続されるインタフェースを含む。更に、各ＤＡＳＤはインタフェースから受信されるデータを選択的に記憶する１つ以上のバッファ、及び記憶装置を含む。各ＤＡＳＤは更に、制御装置からのコマンドに応答して、バッファに受信されたデータ・ブロックを選択し、選択ブロックに対してＸＯＲ演算を実行するプロセッサを含む。プロセッサはまた、その演算結果をバスを介して、別のＤＡＳＤまたは制御装置などの宛先に転送しうる。

【００２６】本発明のデータ記憶システムは、複数ＸＯＲ演算の実行を要求するタスクを、それらＸＯＲ演算の実行をＤＡＳＤ間で分散して実行する。より詳細には、データ記憶システムがＤＡＳＤ間で、データを選択"デイジー・チェイン"順に順次的に転送するように動作する。複数ＤＡＳＤは、データをデイジー・チェイン内の次のＤＡＳＤに転送する以前に、データに対してＸＯＲ演算を実行する。

【００２７】従って、本発明はそのユーザに複数の別個の利点を提供する。第１に、本発明は以前の構成に勝る信頼性を提供する。更に、本発明は、既知の構成に勝る計算効率を提供する。本発明はまた、制御装置の作業負荷を軽減し、ＤＡＳＤ間で処理作業負荷を分散することにより、任意のＤＡＳＤの過剰作業を回避する。

【００２８】

【発明の実施の形態】一般に、本発明は制御装置及び複数のＤＡＳＤを含むデータ記憶装置を含み、ＤＡＳＤは制御装置の計算作業負荷を軽減するために、それらの間でデータを処理し、順次転送することができ、それにより計算効率及びデータ信頼性の向上を達成する。

15

【００２９】ハードウェア要素及び相互接続：図４は本発明の好適なデータ記憶システム４００を示す。システム４００は好適にはＲＡＩＤ－５プロトコルを実現するが、ＲＡＩＤ－１、ＲＡＩＤ３または別のＲＡＩＤ構成が使用されてもよい。システム４００は制御装置４０２を含み、これはＤＡＳＤ４０４乃至４０７に電気的に接続される。図示の態様では、制御装置４０２は、本明細書で述べられるオペレーションを実行する個別の回路要素または特定のアプリケーション向け集積回路（ＡＳＩＣ）を含む。代わりに、制御装置４０２が、本明細書で述べられる機能を達成するために好適なソフトウェアを実行するデジタル・コンピュータ・システムを含んでもよい。

【００３０】好適な実施例では、ＤＡＳＤ４０４乃至４０７及び制御装置４０２は、バス４１０により接続される。より詳細には、制御装置４０２はインタフェース４１１を含み、各ＤＡＳＤ４０４乃至４０７はインタフェース４１２乃至４１５をそれぞれ含む。インタフェース４１２乃至４１５は、好適には直列記憶アーキテクチャなどの、ＳＣＳＩコマンド・セットと互換の直列インタフェースを含む。バス４１０はインタフェース４１１乃至４１５を１連続ループに相互接続し、各インタフェースは２つの隣接インタフェース間に接続される。情報を２重ＤＡＳＤインタフェース（図示せず）間で伝達したり、バス４１０が故障の場合のバックアップを提供するために、２重バス（図示せず）も提供されうる。

【００３１】制御装置４０２及びＤＡＳＤ４０４乃至４０７を相互接続する代わりに、他の構成も使用されうる。例えば制御装置４０２及びＤＡＳＤ４０４乃至４０７が共通バスに並列に接続されたり、ファイバ・チャネル・アービトレーテッド・ループ・タイプの相互接続が使用されてもよい。或いは、制御装置と各ＤＡＳＤ４０４乃至４０７間、及び任意のＤＡＳＤと別のＤＡＳＤ間で信号を交換する特定の手段を提供する、異なる相互接続機構が使用されてもよい。

【００３２】図示の態様では、各ＤＡＳＤ４０４乃至４０７がインタフェース、プロセッサ、バッファ及び記憶装置を含む。例えば図４でＤＡＳＤ４０４はインタフェース４１２、プロセッサ４１８、データ記憶装置４２０及びバッファ４２２を含む。典型的なＤＡＳＤ５００の構成要素が、図５に詳細に示される。ＤＡＳＤ５００はインタフェース５０２、プロセッサ５０４、データ記憶装置５０６、及びバッファ５０８を含む。インタフェース５０２は、第１及び第２のバッファ５１４、５１６を含み、これらはそれぞれバス４１０の第１の終端５１０及び第２の終端５１２に電気的に接続される。インタフェース５０２はまたゲート５１８を含み、これはバス４１０の終端５１０、５１２間で信号を選択的に交換し、更にプロセッサ５０４がバス４１０からの受信データにオペレーションを実行することを支援する。

16

【００３３】プロセッサ５０４はバッファ５１４、５１６に電気的に接続され、それらとデータを交換する。プロセッサ５０４はマイクロプロセッサ、個別論理回路、またはＡＳＩＣを含み、データを処理し本発明に従いＤＡＳＤ５００のオペレーションを管理する。図示の態様では、プロセッサ５０４は、例えばモデル８０１８６マイクロプロセッサ、またはウェスタン・デジタル社のモデルＷＤ３３Ｃ９６及びＷＤ６１Ｃ４０回路の組合わせを含みうる。プロセッサ５０４は、記憶装置５０６及びバッファ５０８にも電気的に接続される。

【００３４】図示の態様では、記憶装置５０６は、１つ以上の回転磁気記録媒体を有するハード・ディスク・ドライブを含みうる。或いは、ＤＡＳＤ４０４乃至４０７が光ディスク・ドライブ、磁気テープ・ドライブ、フロッピーまたはリジッド・ディスケット、ランダム・アクセス・メモリ（ＲＡＭ）、または他の好適なコンピュータ読出し可能データ記憶装置を含んでもよい。バッファ５０８、５１４及び５１６は、好適にはＤＲＡＭバッファを含み、バッファ５０８は後述される理由により、複数の従属バッファに区分化されることが望ましい。

【００３５】オペレーション方法：上述のデータ記憶システムの様々な態様に加え、本発明はこうしたシステムをオペレートする方法を包含する。一般に、本発明のデータ記憶システムをオペレートするために、制御装置４０２（図４）はコマンド及びデータをバス４１０を介して、ＤＡＳＤ４０４乃至４０７に送信する。ＤＡＳＤ４０４乃至４０７は、データ記憶システム全体の読出し、書込み、再生、及び他のオペレーションの実行を支援するために要求される様々な局所タスクを実行して応答する。本発明の図示の態様の利点は、ＸＯＲ演算がＤＡＳＤ４０４乃至４０７間で分散され、ＸＯＲ演算を要求する所与の機能に対して、どのＤＡＳＤも複数のＸＯＲ演算を実行するように要求されず、従って処理作業負荷の過度な分配を請け負わないことである。

【００３６】コマンド・セット：本発明の図示の態様では、制御装置４０２は様々なコマンドを選択し、ＤＡＳＤ４０４乃至４０７に発行することにより、様々な特定のオペレーション（後述）を実行する。本発明の図示の態様では、各ＤＡＳＤがそのゲート５１８により、プロセッサ５０４とバッファ５１４、５１６との間で信号を選択的に交換しうる。例えば、制御装置４０２がＤＡＳＤ４０６に対してコマンドを発行する場合、制御装置４０２はそのコマンドをバス４１０上に出力する。コマンドはインタフェース４１２、４１３及び４１４を介して、インタフェース４１５に転送される。この状況では、各インタフェース４１２、４１３、４１４のゲート５１８は、コマンドが別のＤＡＳＤに向けられることを認識し、信号をバッファ５１４からバッファ５１６に転送し、事実上、バス４１０の終端５１０、５１２を接続する。

【００３７】各コマンドは、ＤＡＳＤ４０４乃至４０７により認識される全てのコマンドを含む"コマンド・セット"から選択される。本発明の図示の態様では、コマンド・セットが次のコマンドを含みうる。ＷＲＩＴＥ、ＸＴ、ＸＷ、ＸＴＷ、ＸＴＷ／ＮＲ、ＴＷ、ＲＥＡＤ、及びＴＲＡＮＳＦＥＲ。

【００３８】ＷＲＩＴＥ：このコマンドは、ターゲットＤＡＳＤ５００に、特定の入来データをその記憶装置５０６に記憶するように指令する。この点に関し、"入来"データは、インタフェース５０２によりバス４１０から獲得されるデータを指す。従って、入来データは制御装置４０２、別のＤＡＳＤ、または別の制御装置（図示せず）などの発信元から発信されうる。入来データが制御装置４０２以外の発信元から提供される場合、制御装置４０２もまたＴＲＡＮＳＦＥＲコマンド（後述）をデータ発信元に送信する。

【００３９】ＸＴ：このコマンドは、ターゲットＤＡＳＤ５００に、特定のデータをその記憶装置５０６から読出し、そのデータと特定の入来データの組合わせにＸＯＲを実行し、ＸＯＲ結果を２次ターゲットに転送するように指令する。ここでの説明の都合上、"２次ターゲット"は別のＤＡＳＤ、制御装置４０２、または異なる制御装置（図示せず）を含みうるものとする。ＷＲＩＴＥコマンド同様、入来データは制御装置４０２、別のＤＡＳＤ、または別の制御装置（図示せず）などの発信元から発信されうる。入来データが制御装置４０２以外の発信元から提供される場合、制御装置４０２もまたＴＲＡＮＳＦＥＲコマンドをデータ発信元に送信する。

【００４０】ターゲットＤＡＳＤ５００がＸＴコマンドを受信すると、インタフェース５０２が入来データを受信し、記憶装置５０６がＸＴコマンドにより指定されるデータを読出す。これらのデータ項目は、好適には続く処理のためにバッファ５０８に記憶される。特に、プロセッサ５０４は、バッファ５０８内に含まれる読出されたデータ及び入来データに対して、ＸＯＲ演算を実行する。次にターゲットＤＡＳＤ５００が２次イニシエータとして、限定的役割を請け負う。特に、プロセッサ５０４はＴＲＡＮＳＦＥＲコマンド（後述）を発行して、ＸＯＲ結果をバス４１０を介して、指定２次ターゲットに転送する。

【００４１】ターゲットＤＡＳＤ５００は、実現されるＤＡＳＤの特定の態様に従い、多彩な方法により２次ターゲットを識別しうる。１つの態様では、ターゲットＤＡＳＤ５００はＤＡＳＤ５００に局所的に記憶されるテーブルを調査し、ＸＯＲ結果の送り先を決定する。このテーブルは、ターゲットＤＡＳＤ５００と他のＤＡＳＤとの関係、制御装置４０２から受信されるコマンド、及びＸＯＲデータの発信元を考慮する。この情報は、ターゲットＤＡＳＤ５００が適切な２次ターゲットを選択するために使用される。別の態様では、制御装置４０２からのコマンドが、ターゲットＤＡＳＤ５００に情報の転送先を指令するパラメータを含む。上述の両方の態様は本開示を参考に、当業者により実現されよう。

【００４２】ＸＷ：このコマンドはターゲットＤＡＳＤ５００に、特定の入来データ及びその記憶装置５０６からの特定のデータに対して、ＸＯＲ演算を実行し、ＸＯＲ結果を記憶装置に書込むように指令する。ＸＴコマンド同様、入来データは制御装置４０２、別のＤＡＳＤ、または別の制御装置（図示せず）などの発信元から発信されうる。

【００４３】ターゲットＤＡＳＤ５００がＸＷコマンドを受信すると、インタフェース５０２が入来データを受信し、記憶装置５０６がコマンドにより指定されるデータを読出す。これらのデータ項目は、好適には続く処理のためにバッファ５０８に記憶される。次に、プロセッサ５０４が、バッファ５０８に含まれる２つのデータ項目に対して、ＸＯＲ演算を実行し、記憶装置５０６がＸＯＲ結果を記憶する。

【００４４】ＸＴＷ：このコマンドはターゲットＤＡＳＤ５００に、特定の入来データ及びその記憶装置５０６からの特定のデータに対して、ＸＯＲ演算を実行し、入来データをその記憶装置５０６に書込み、ＸＯＲ結果を２次ターゲットに転送するように指令する。インタフェース５０２はバス４１０を介して入来データを受信し、入来データは制御装置４０２、別のＤＡＳＤ、または別の発信元（図示せず）から発信されうる。

【００４５】ターゲットＤＡＳＤ５００がＸＴＷコマンドを受信すると、記憶装置５０６がコマンドにより指定されるデータを読出し、インタフェース５０２が入来データを受信し、プロセッサ５０４が２つのデータ項目に対して、ＸＯＲ演算を実行する。好適には、これら２つのデータ項目はプロセッサ５０４を支援するために、バッファ５０８に記憶される。次に、記憶装置５０６が入来データを記憶する。次にターゲットＤＡＳＤ５００が２次イニシエータとして、限定的役割を請け負い、プロセッサ５０４はＴＲＡＮＳＦＥＲコマンド（後述）を発行して、ＸＯＲ結果をインタフェース５０２及びバス４１０を介して、指定２次ターゲットに転送する。

【００４６】ＸＴＷ／ＮＲ：このコマンドはターゲットＤＡＳＤ５００に、入来データの旧ブロック及び新ブロックに対してＸＯＲ演算を実行し、新データ・ブロックをターゲットＤＡＳＤの記憶装置に書込み、ＸＯＲ結果を２次ターゲットに転送するように指令する。新たな入来データ・ブロックは制御装置４０２から発信され、旧データ・ブロックは制御装置４０２または別のＤＡＳＤから発信されうる。ターゲットＤＡＳＤはバス４１０を介して、入来データの新旧両ブロックを受信する。

【００４７】ターゲットＤＡＳＤ５００がＸＴＷ／ＮＲコマンドを受信すると、インタフェース５０２が２つの入来データ・ブロックを受信し、バッファ５０８が好適

にはデータ・ブロックを記憶する。次にプロセッサ５０４が２つのデータ・ブロックに対してＸＯＲ演算を実行し、記憶装置５０６が新たなデータ・ブロックを記憶する。２つのデータ・ブロックが制御装置４０２から受信される場合、新データ・ブロックは、例えばターゲットＤＡＳＤ５００に転送される最初のデータ・ブロックであることから識別される。ターゲットＤＡＳＤ５００が制御装置４０２から新データ・ブロックを受信し、旧データ・ブロックを別のＤＡＳＤから受信する場合、ターゲットＤＡＳＤの記憶装置５０６は、制御装置４０２から受信されるデータ・ブロック（すなわち新データ・ブロック）を記憶する。

【００４８】記憶装置５０６が入来データから新データ・ブロックを記憶した後、ターゲットＤＡＳＤ５００は２次イニシエータとして、限定的役割を請け負う。プロセッサ５０４はＴＲＡＮＳＦＥＲコマンドを発行して、ＸＯＲ結果をインタフェース５０２及びバス４１０を介して、指定２次ターゲットに転送する。

【００４９】ＴＷ：このコマンドはＤＡＳＤ５００に、入来データをその記憶装置５０６に書込み、また入来データを２次ターゲットに転送するように指令する。特に、ＴＷコマンドに応答して、インタフェース５０２が入来データを受信し、記憶装置５０６が入来データを記憶し、ＤＡＳＤ５００が２次イニシエータとして限定的役割を請け負う。プロセッサ５０４はＴＲＡＮＳＦＥＲコマンドを発行し、データを２次ターゲットに転送する。

【００５０】ＲＥＡＤ：このコマンドはターゲットＤＡＳＤ５００に、その記憶装置５０６の指定アドレスからデータを読出し、読出したデータをバス４１０を介して２次ターゲットに転送するように指令する。ＲＥＡＤコマンドは、データを読出す記憶装置５０６のアドレスを指定する。記憶装置５０６から読出されたデータは、図示の態様では、バッファ５０８に転送され、次にインタフェース５０２を介して、２次ターゲットに転送される。

【００５１】ＴＲＡＮＳＦＥＲ：このコマンドはデータ発信元により発行され、ターゲットＤＡＳＤ５００に、データ発信元からのデータをその記憶装置５０６に受信するように準備させる。２次ターゲットが転送データを受信すると、２次ターゲットは、以前に制御装置４０２により指令されたオペレーションを完了することができる。より詳細には、ＴＲＡＮＳＦＥＲコマンドは第１のＤＡＳＤにより、制御装置４０２により以前に発行されたコマンドを完了するために使用され、それにより、第２のＤＡＳＤが第１のＤＡＳＤからデータを受信することを支援する。特に、この態様では、第１のＤＡＳＤがＴＲＡＮＳＦＥＲコマンドを第２のＤＡＳＤに送信し、データが第１のＤＡＳＤから第２のＤＡＳＤへ転送されることを事前に警告する。

【００５２】包括的（comprehensive）マスタ・オペレーション：上述のコマンド・セットは、特定の結果を獲得するために、個々のＤＡＳＤに転送される個々のコマンドのグループを含む。より重要な点は、これらのコマンドの内の選択コマンドを共通体系として複数ＤＡＳＤに発行することにより、制御装置４０２がＤＡＳＤ内で包括的マスタ・オペレーションを開始することができ、実際に最小の作業負荷が制御装置４０２により実行されることである。本発明の重要な特長の１つは、以降で詳述されるように、特定のマスタ・オペレーションがＤＡＳＤ４０４乃至４０７を通じて、データの"デイジー・チェイン"転送を有利に使用しうることである。デイジー・チェイン・データ転送により、データは各ＤＡＳＤを通じて直列に渡される。すなわち、イニシエータＤＡＳＤから開始し、複数の中間ＤＡＳＤを通過し、最終的に受信側ＤＡＳＤに到来する。好適には、データはデイジー・チェイン内の各ＤＡＳＤを通じて１度だけ転送される。後述されるように、特定のケースでは、ＤＡＳＤがデータをその転送以前に処理し、また他のケースでは、ＤＡＳＤがデータを処理すること無く、単に転送する。すなわち、コンジット（conduit）として機能する。この処理の例として、各中間ＤＡＳＤが、デイジー・チェイン内の直前のＤＡＳＤから受信されるデータ、並びに制御装置４０２またはその中間ＤＡＳＤの記憶装置から受信される別のデータ項目に対して、ＸＯＲ演算を実行する。このように、本発明のデイジー・チェイン転送はＤＡＳＤ間のＸＯＲタスクについて述べ、それにより制御装置４０２の負荷を軽減し、どのＤＡＳＤも不均等な負荷を負わないことを保証する。

【００５３】図６乃至図１１に示されるように、これらの包括的マスタ・オペレーションには、UPDATE、REGENERATE、REBUILD、UPDATE（CACHED DATA）、UPDATE（FAILED DASD）、及びWRITE STRIPEが含まれる。

【００５４】UPDATE（更新）：このオペレーションは一般に、新たなデータ・ブロックをターゲットＤＡＳＤに書込み、新たなデータに対応するパリティ・ビットを更新する。更新書込み（UPDATE WRITE）オペレーションについて、図６を参照して詳細に説明しよう。図６は、制御装置６０１及びＤＡＳＤ６０２乃至６０５を含むデータ記憶システム６００を示す。ＤＡＳＤ６０５はパリティＤＡＳＤとして指定される。しかしながら、各ＤＡＳＤの特定部分を、他のディスクＤＡＳＤに対応するパリティ・ビットを記憶するように割当ててもよい。ここでは説明の都合上、システム６００が、パリティ情報を専用に記憶する特定のＤＡＳＤ６０５を有するものとする。

【００５５】制御装置６０１は、ＸＴＷコマンド６０８及び新たなデータ・ブロック６１０をＤＡＳＤ６０２に、またＸＷコマンド６０９をＤＡＳＤ６０５に、次のシーケンスの実行に適切なパラメータと一緒に送信する

ことにより、更新オペレーションを開始する。上述のＸＴＷコマンドに従い、ＤＡＳＤ６０２が新たなデータ６１０及びその記憶装置からの対応データに対して、ＸＯＲ演算を実行し、ＸＯＲ結果６１５を生成する。次にＤＡＳＤ６０２は新たなデータ６１０をその記憶装置に書込み、ＸＯＲ結果をパリティＤＡＳＤ６０５に転送する。パリティＤＡＳＤ６０５は受信ＸＷコマンド６０９に従い、受信ＸＯＲ結果６１５及びその記憶装置からの読出しデータに対してＸＯＲ演算を実行し、このＸＯＲ演算結果をその記憶装置に書込む。

【００５６】このように、新たなデータ・ブロック６１０がＤＡＳＤ６０２に書込まれ、対応パリティがパリティＤＡＳＤ６０５内で更新される。

【００５７】REGENERATE（再生）：REGENERATEコマンドが図７に示され、ここでデータ記憶システム７００は制御装置７０１及びＤＡＳＤ７０２乃至７０６を含み、ＤＡＳＤ７０６は故障している。制御装置７０１は、ＲＥＡＤコマンド７０８をＤＡＳＤ７０２に、またＸＴコマンド７１０乃至７１２をＤＡＳＤ７０３乃至７０５に、次のシーケンスの実行に適切なパラメータと一緒に送信することにより、再生オペレーションを開始する。上述のＲＥＡＤコマンドに従い、ＤＡＳＤ７０２がコマンド７０８により要求されるデータ・ブロックをその記憶装置から読出し、それをＤＡＳＤ７０３に転送する。各ＤＡＳＤ７０３乃至７０４が別のＤＡＳＤからデータ・ブロックを受信すると、それらはそのデータ・ブロック及びその記憶装置からの対応データに対してＸＯＲ演算を実行し、ＸＯＲ結果を次のＤＡＳＤに順番に転送する。ＤＡＳＤ７０５の場合には、ＸＯＲ結果は制御装置７０１に返送される。コマンド７０８、７１０乃至７１２に付随して制御装置７０１により発行されるパラメータは、ＤＡＳＤ７０２乃至７０５間のデータ転送の順序を決定する（例えばＤＡＳＤ７０２からＤＡＳＤ７０３へ、次にＤＡＳＤ７０４へなど）。

【００５８】このように、故障ＤＡＳＤ７０６の所望のロケーションからのデータが、ＤＡＳＤ７０２乃至７０５内のデイジー・チェインＸＯＲ計算により復元され、制御装置７０１に提供される。

【００５９】REBUILD（復元）：REBUILDコマンドが図８に示され、ここでデータ記憶システム８００は制御装置８０１及びＤＡＳＤ８０２乃至８０５を含む。ＤＡＳＤ８０５は、復元されなければならないデータを有するドライブとして指定される。制御装置８０１は、ＲＥＡＤコマンド８０８をＤＡＳＤ８０２に送信し、ＸＴコマンド８１０乃至８１１をＤＡＳＤ８０３乃至８０４に送信し、ＷＲＩＴＥコマンド８１４をＤＡＳＤ８０５に送信することにより、復元オペレーションを開始する。全てのコマンドは、続くシーケンスの実行に適切なパラメータを添付される。上述のＲＥＡＤコマンドに従い、ＤＡＳＤ８０２はＲＥＡＤコマンド８０８により要求されるデータを、その記憶装置から読出し、それをＤＡＳＤ８０３に転送する。ＤＡＳＤ８０３がデータ・ブロックをＤＡＳＤ８０２から受信すると、これはこのデータ・ブロック及びその記憶装置からの対応データにＸＯＲ演算を実行し、ＸＯＲ結果を次のＤＡＳＤ８０４に順番に転送する。ＤＡＳＤ８０４はＤＡＳＤ８０３と同様なオペレーションを実行する。ＤＡＳＤ８０５は単にＸＯＲ結果をＤＡＳＤ８０４から受信し、その結果を自身の記憶装置（すなわちＤＡＳＤ８０５の記憶装置）に記憶する。

【００６０】このように、ＤＡＳＤ８０５の所望のロケーションからのデータが、ＤＡＳＤ８０３乃至８０４のデイジー・チェインＸＯＲ計算により復元され、ＤＡＳＤ８０５の適切なロケーションに書き戻される。上述の再生オペレーションと同様、コマンド８０８、８１０乃至８１１及び８１４に付随して、制御装置８０１により発行されるパラメータが、ＤＡＳＤ８０２乃至８０５間のデータ転送の順序を決定する。

【００６１】UPDATE(CACHED DATA)（キャッシュ・データによる更新）：UPDATE(CACHED DATA)コマンドが図９に示され、ここでデータ記憶システム９００は、制御装置９０１及びＤＡＳＤ９０２乃至９０５を含む。ＤＡＳＤ９０５はパリティＤＡＳＤとして指定される。しかしながら、各ＤＡＳＤの特定部分を、他のＤＡＳＤに対応するパリティ・ビットを記憶するように割当ててもよい。ここでは説明の都合上、システム９００が、パリティ情報を専用に記憶する特定のＤＡＳＤ９０５を有するものとする。

【００６２】キャッシュ・データによる更新オペレーションは、一般に、新たなデータ・ブロックをターゲットＤＡＳＤ９０２に記憶することにより、ターゲットＤＡＳＤを更新し、それに従いパリティＤＡＳＤ９０５を更新する。しかしながら、図６の更新オペレーションと異なり、新たなデータ・ブロックにより置換される旧データ・ブロックが、以前に制御装置９０１にキャッシュされている。従って、ターゲットＤＡＳＤ９０２は旧データをその記憶装置から読出す必要はなく、オペレーションはより迅速に実行される。

【００６３】制御装置９０１は、ＸＴＷ／ＮＲコマンド９０８、新たなデータ・ブロック９１２、及び旧データ・ブロック９１３をＤＡＳＤ９０２に送信することにより、キャッシュ・データによる更新オペレーションを開始する。制御装置９０１はＸＷコマンド９１０をパリティＤＡＳＤ９０５に送信する。これらのコマンドは、続くシーケンスの実行に適切なパラメータと一緒に送信される。上述のＸＴＷ／ＮＲコマンドに従い、ＤＡＳＤ９０２がデータ・ブロック９１２乃至９１３に対してＸＯＲ演算を実行し、新たなデータ・ブロック９１２をＤＡＳＤ９０２の記憶装置に書込み、ＸＯＲ結果をパリティＤＡＳＤ９０５に転送する。パリティＤＡＳＤ９０５は

ＸＷコマンド（上述）９１０に従い、受信ＸＯＲ結果及びその記憶装置からの対応データに対してＸＯＲ演算を実行し、最終ＸＯＲ結果を計算する。次にパリティＤＡＳＤ９０５が最終ＸＯＲ結果を、その記憶装置の対応ロケーションに記憶する。

【００６４】このように、新たなデータ・ブロック９１２がＤＡＳＤ９０２に書込まれ、対応パリティがパリティＤＡＳＤ９０５内で更新される。

【００６５】UPDATE(FAILED DASD)（故障ＤＡＳＤ状況の更新）：このオペレーションは一般に、ＤＡＳＤの故障のために更新できないデータ・ブロックに対応するパリティ・ビットを更新する。図１０を参照して、故障ＤＡＳＤ状況の更新オペレーションをより詳しく説明しよう。図１０において、データ記憶システム１０００は、制御装置１００１及びＤＡＳＤ１００２乃至１００５を含む。ＤＡＳＤ１００５はパリティＤＡＳＤとして指定される。しかしながら、各ＤＡＳＤの特定部分を、他のＤＡＳＤに対応するパリティ・ビットを記憶するように割当ててもよい。ここでは説明の都合上、システム１０００が、パリティ情報を専用に記憶する特定のＤＡＳＤ１００５を有するものとする。

【００６６】制御装置１００１は、ＸＴコマンド１００８及び新データ・ブロック１００９をターゲットＤＡＳＤ１００２に送信し、ＸＴコマンド１０１０乃至１０１１をＤＡＳＤ１００３乃至１００４に送信し、ＷＲＩＴＥコマンド１０１２をＤＡＳＤ１００５に送信することにより、故障ＤＡＳＤ状況の更新オペレーションを開始する。全てのコマンドは、続くシーケンスの実行に適切なパラメータを付随される。上述のＸＴコマンドに従い、最初のＤＡＳＤ１００２は、新たなデータ１００９及びその記憶装置からの対応データに対してＸＯＲ演算を実行し、ＸＯＲ結果をＤＡＳＤ１００３に転送する。

【００６７】同様に、ＤＡＳＤ１００３乃至１００４は、それぞれＤＡＳＤ１００２、１００３から受信されるデータ、及びそれらの記憶装置からの対応データに対して、ＸＯＲ演算を実行する。これらのＸＯＲ結果は、図１０に示されるように、それぞれＤＡＳＤ１００４及び１００５に転送される。上述のＷＲＩＴＥコマンドに従い、パリティＤＡＳＤ１００５がＤＡＳＤ１００４からＸＯＲ結果を受信し、それをＤＡＳＤ１００５の記憶装置に書込むことにより、この結果を記憶する。

【００６８】このように、新たなデータ・ブロック１００９が故障ＤＡＳＤ１００６に書込まれないものの、ＤＡＳＤ１００２乃至１００４のデイジー・チェインＸＯＲ計算により、パリティＤＡＳＤ１００５が更新される。

【００６９】WRITE STRIPE（ストライプ書込み）：このオペレーションは一般に、新データ・ブロックの"ストライプ"をＤＡＳＤに書込み、新データ・ブロックに対応するパリティ・ビットを更新する。図１１を参照して、ストライプ書込みオペレーションを詳細に説明しよう。データ記憶システム１１００は、制御装置１１０１及びＤＡＳＤ１１０２乃至１１０５を含み、ＤＡＳＤ１１０５はパリティＤＡＳＤとして指定される。しかしながら、各ＤＡＳＤの特定部分を、他のＤＡＳＤに対応するパリティ・ビットを記憶するように割当ててもよい。ここでは説明の都合上、システム１１００が、パリティ情報を専用に記憶する特定のＤＡＳＤ１１０５を有するものとする。

【００７０】制御装置１１０１は、ＴＷコマンド１１０８及び新データ・ブロック１１０９をＤＡＳＤ１１０２に送信し、ＸＴＷ／ＮＲコマンド１１１０及び新データ・ブロック１１１１をＤＡＳＤ１１０３に送信し、ＸＴＷ／ＮＲコマンド１１１２及び新データ・ブロック１１１３をＤＡＳＤ１１０４に送信し、ＷＲＩＴＥコマンド１１１４をＤＡＳＤ１１０５に送信することにより、ストライプ書込みオペレーションを開始する。これらのコマンドは、続くシーケンスの実行に適切なパラメータと一緒に送信される。上述のＴＷコマンドに従い、ＤＡＳＤ１１０２はデータ・ブロック１１０９を受信し、それをＤＡＳＤ１１０２に対応する記憶装置に記憶する。ＤＡＳＤ１１０２は更に受信データ・ブロック１１０９をＤＡＳＤ１１０３に転送する。上述のＸＴＷ／ＮＲコマンドに従い、ＤＡＳＤ１１０３は、新データ・ブロック１１１１及びＤＡＳＤ１１０２から転送されるデータ・ブロックを受信し、これら２つのデータ・ブロックに対してＸＯＲ演算を実行する。ＤＡＳＤ１１０３もまた新データ・ブロック１１１１をその記憶装置に書込み、そのＸＯＲ結果をＤＡＳＤ１１０４に転送する。同様に、ＤＡＳＤ１１０４は、新データ・ブロック１１１３及びＤＡＳＤ１１０３から転送されるデータ・ブロックを受信し、これら２つのデータ・ブロックに対してＸＯＲ演算を実行する。ＤＡＳＤ１１０４もまた新データ・ブロック１１１３をその記憶装置に書込み、そのＸＯＲ結果をＤＡＳＤ１１０５に転送する。ＷＲＩＴＥコマンド１１１４に従い、パリティＤＡＳＤ１１０５がＤＡＳＤ１１０４から受信されるＸＯＲ結果を、パリティＤＡＳＤ１１０５の記憶装置に書込む。

【００７１】このように、各データ・ブロックが対応ＤＡＳＤドライブ１１０２乃至１１０４に書込まれ、対応パリティ・ビットがＤＡＳＤ１１０２乃至１１０４のデイジー・チェインＸＯＲ計算により更新され、更新パリティ・ビットがパリティＤＡＳＤ１１０５に記憶される。

【００７２】他の包括的マスタオペレーション：上述の包括的マスタ・オペレーションに加え、ここで述べられるコマンド・セットが、本発明のデータ記憶システムと一緒に、様々な他のオペレーションを実行するために使用されうる。更に、新たなコマンド及びここで述べられるコマンドの適応が、本発明の範囲内で、データ記憶シ

ステムにおいて実現されうる。こうしたオペレーション及びコマンドは、当業者の能力の範囲内で達成されよう。また、ここでは本発明の特定の実施例についてのみ述べてきたが、当業者には本発明の範囲から逸脱すること無しに、様々な変更及び他の実施例が実現可能であることが理解されよう。

【００７３】まとめとして、本発明の構成に関して以下の事項を開示する。

【００７４】（１）マシン読出し可能データを記憶する装置で使用されるＤＡＳＤであって、前記装置が制御装置、複数のＤＡＳＤ、及び前記制御装置と前記ＤＡＳＤを電気的に相互接続するバスを含み、前記複数のＤＡＳＤのうちの１つが、前記制御装置の各オペレーションにおけるパリティＤＡＳＤとして指定されるものにおいて、マシン読出し可能データを記憶する記憶装置と、前記バスに電気的に接続されるインタフェースと、前記インタフェース及び前記記憶装置に動作的に接続される少なくとも１つのバッファと、前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介してコマンドを受信し、該コマンドに応答して、前記バッファ内に記憶されるデータ項目を選択し、前記選択データ項目に対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を前記バスを介して、前記コマンドにより指定される、前記ＤＡＳＤの１つまたは前記制御装置を含む宛先に送信するプロセッサと、を含む、ＤＡＳＤ。

（２）前記プロセッサがマイクロプロセッサを含む、上記（１）記載のＤＡＳＤ。

（３）前記プロセッサがＡＳＩＣを含む、上記（１）記載のＤＡＳＤ。

（４）前記プロセッサが個別回路要素のアセンブリから成る論理回路を含む、上記（１）記載のＤＡＳＤ。

（５）前記記憶装置が、少なくとも１つの磁気記録ディスクを含むディスク・ドライブを含む、上記（１）記載のＤＡＳＤ。

（６）前記インタフェースが直列インタフェースを含む、上記（１）記載のＤＡＳＤ。

（７）前記インタフェースが並列インタフェースを含む、上記（１）記載のＤＡＳＤ。

（８）バスにより相互接続される複数のメンバを含むデータ記憶システムにおいて使用されるＤＡＳＤであって、前記メンバが前記制御装置及び前記複数のＤＡＳＤを含むものにおいて、前記バスに電気的に接続されるインタフェースと、データ記憶装置と、前記インタフェース及び前記記憶装置に動作的に接続され、それらから受信されるデータを選択的に記憶するバッファと、前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介して受信されるコマンドに応答して、次のステップ、すなわち、前記バッファと前記インタフェース間でデータを選択的に転送するステップ、前記記憶装置と前記バッファ間でデータを交換するステップ、前記バッファ内に含まれる前記選択データ項目に対してＸＯＲ演算を実行するステップ、及び前記インタフェースから選択される前記メンバにデータを転送するステップの少なくとも１つを実行するようにプログラムされるプロセッサと、を含む、ＤＡＳＤ。

（９）前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記記憶装置から選択データを読出し、該選択データを選択される前記メンバに転送するステップを実行するようにプログラムされる、上記（８）記載のＤＡＳＤ。

（１０）前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記インタフェースにより受信される前記選択データを前記記憶装置に書込むステップを実行するようにプログラムされる、上記（８）記載のＤＡＳＤ。

（１１）前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記記憶装置から選択データを読出し、前記インタフェースにより受信されるデータ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送するステップを実行するようにプログラムされる、上記（８）記載のＤＡＳＤ。

（１２）前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記記憶装置から前記選択データを読出し、前記インタフェースにより受信されるデータ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送し、前記インタフェースにより受信される前記データを前記記憶装置に書込むステップを実行するようにプログラムされる、上記（８）記載のＤＡＳＤ。

（１３）前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記記憶装置から前記選択データを読出し、前記インタフェースにより受信されるデータ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を前記記憶装置に書込むステップを実行するようにプログラムされる、上記（８）記載のＤＡＳＤ。

（１４）前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記インタフェースにより受信される前記選択データを前記記憶装置に書込み、前記選択データを選択される前記メンバに転送するステップを実行するようにプログラムされる、上記（８）記載のＤＡＳＤ。

（１５）前記プロセッサが、前記制御装置からの少なくとも１つの所定のコマンドに応答して、前記インタフェースにより受信される２つのデータ項目に対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送し、選択される一方の前記受信データ項目を前記記憶装置に書込むステップを実行するようにプログラムされる、上記（８）記載のＤＡＳＤ。

（１６）マシン読出し可能データを記憶するシステムであって、制御装置及び複数のＤＡＳＤを含む複数のメンバと、前記メンバを電気的に相互接続してループを形成するバスと、を含み、前記の各ＤＡＳＤが、マシン読出し可能データを記憶する記憶装置と、前記バスに電気的に接続されるインタフェースと、前記インタフェース及び前記記憶装置に動作的に接続され、前記インタフェース及び前記記憶装置から受信されるデータを選択的に記憶する少なくとも１つのバッファと、前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介してコマンドを受信し、所定のコマンドに応答して、前記バッファ内に記憶される前記データ項目を選択し、前記選択データ項目に対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を前記バスを介して、選択される前記メンバを含む宛先に送信するＸＯＲ発生器と、を含む、システム。

（１７）前記プロセッサがマイクロプロセッサを含む、上記（１６）記載のシステム。

（１８）前記プロセッサがＡＳＩＣを含む、上記（１６）記載のシステム。

（１９）前記プロセッサが個別回路要素のアセンブリから成る論理回路を含む、上記（１６）記載のシステム。

（２０）前記記憶装置が、少なくとも１つの磁気記録ディスクを含むディスク・ドライブを含む、上記（１６）記載のシステム。

（２１）前記インタフェースが直列インタフェースを含む、上記（１６）記載のシステム。

（２２）前記インタフェースが並列インタフェースを含む、上記（１６）記載のシステム。

（２３）前記複数のＤＡＳＤがＲＡＩＤプロトコルに従い構成される、上記（１６）記載のシステム。

（２４）前記ＲＡＩＤプロトコルがＲＡＩＤ－５を含む、上記（２３）記載のシステム。

（２５）複数のＸＯＲ演算の実行を要求するタスクを実行するように構成されるデータ記憶システムであって、コマンドを転送する制御装置と、前記コマンドに応答して、前記ＸＯＲ演算の実行をＤＡＳＤ間で分散することにより、前記タスクを実行するＤＡＳＤアレイと、前記ＤＡＳＤ及び前記制御装置を相互接続するバスと、を含む、データ記憶システム。

（２６）データ記憶システムを動作させる方法であって、制御装置及び複数のＤＡＳＤのストリングを含む複数のメンバと、前記複数のメンバを電気的に相互接続するバスと、を含み、各ＤＡＳＤが、前記バスに電気的に接続されるインタフェースと、データ記憶装置と、前記インタフェース及び前記記憶装置に動作的に接続されるバッファと、前記バッファに動作的に接続されるプロセッサと、を含み、使用可能な前記ＤＡＳＤを識別するステップと、前記制御装置から前記識別されたＤＡＳＤに個々にコマンドを送信するステップと、前記コマンドに応答して、データが前記識別されたＤＡＳＤを所定順序でデイジー・チェイン転送されるように、前記識別されたＤＡＳＤを動作させるステップと、を含み、データがイニシエータＤＡＳＤから始まり、複数の中間ＤＡＳＤを経由してレシーバＤＡＳＤに至る前記各ＤＡＳＤにより、順次転送及び受信され、前記の各中間ＤＡＳＤが、前記デイジー・チェインに沿って受信される前記データを前記所定順序に従う次のＤＡＳＤに転送する以前に、前記受信データ項目及び別のデータ項目に対して、ＸＯＲ演算を実行する、方法。

（２７）前記別のデータ項目が前記制御装置から受信される前記データ項目を含む、上記（２６）記載の方法。

（２８）前記別のデータ項目が前記中間ＤＡＳＤのそれぞれの前記記憶装置から読出される前記データ項目を含む、上記（２６）記載の方法。

（２９）前記制御装置からの少なくとも１つの前記所定コマンドに応答して、選択ＤＡＳＤの前記記憶装置から前記データを読出し、読出したデータを選択される前記メンバに転送するように、前記選択ＤＡＳＤをオペレートするステップを含む、上記（２６）記載の方法。

（３０）前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記インタフェースにより受信される前記データを前記選択ＤＡＳＤの前記記憶装置に書込むように、前記選択ＤＡＳＤをオペレートするステップを含む、上記（２６）記載の方法。

（３１）前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記記憶装置から前記選択データを読出し、前記選択ＤＡＳＤの前記インタフェースにより受信される前記データ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送するように、前記選択ＤＡＳＤをオペレートするステップを含む、上記（２６）記載の方法。

（３２）前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記記憶装置から前記選択データを読出し、前記選択ＤＡＳＤの前記インタフェースにより受信される前記データ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送し、前記選択ＤＡＳＤの前記インタフェースにより受信された前記データを該選択ＤＡＳＤの前記記憶装置に書込むように、前記選択ＤＡＳＤをオペレートするステップを含む、上記（２６）記載の方法。

（３３）前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記記憶装置から前記選択データを読出し、前記選択ＤＡＳＤの前記インタフェースにより受信される前記データ及び前記選択データに対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を前記選択ＤＡＳＤの前記記憶装置に書込むように、前記選択ＤＡＳＤをオペレートするステップを含む、上

記（２６）記載の方法。

（３４）前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記インタフェースにより受信される前記選択データを前記選択ＤＡＳＤの前記記憶装置に書込み、前記受信データを選択される前記メンバに転送するように、前記選択ＤＡＳＤをオペレートするステップを含む、上記（２６）記載の方法。

（３５）前記制御装置からの少なくとも１つの前記所定コマンドに応答して、前記選択ＤＡＳＤの前記インタフェースにより受信される２つの前記データ項目に対してＸＯＲ演算を実行し、該ＸＯＲ演算結果を選択される前記メンバに転送し、選択される一方の前記受信データ項目を前記選択ＤＡＳＤの前記記憶装置に書込むように、前記選択ＤＡＳＤをオペレートするステップを含む、上記（２６）記載の方法。

（３６）制御装置、ＤＡＳＤアレイ、及び前記ＤＡＳＤ及び前記制御装置を相互接続するバスを含むデータ記憶システムにおいて、複数のＸＯＲ演算の実行を要求するタスクを実行するように、前記データ記憶システムを動作させる方法であって、前記制御装置から前記バスを介して前記ＤＡＳＤにコマンドを転送するステップと、前記コマンドに応答して、前記ＸＯＲ演算の実行を前記ＤＡＳＤ間で分散することにより、前記タスクを実行するように、前記ＤＡＳＤをオペレートするステップと、を含む、方法。

（３７）前記オペレーティング・ステップが、データを前記ＤＡＳＤ間で選択順に順次転送するように、前記ＤＡＳＤをオペレートするステップを含み、前記複数のＤＡＳＤが前記選択順序に従い前記データを別のＤＡＳＤに転送する以前に、該データに対して前記ＸＯＲ演算を実行する、上記（３６）記載の方法。

（３８）前記オペレーティング・ステップが再生オペレーションを含み、該再生オペレーションが、第１のＤＡＳＤの第１の記憶ロケーションから第１のデータ・ブロックを読出し、前記第１のデータ・ブロックを第２のＤＡＳＤに転送するように、前記第１のＤＡＳＤをオペレートするステップと、前記第２のＤＡＳＤをオペレートするステップであって、前記第１のデータ・ブロックを前記第１のＤＡＳＤから受信するステップと、前記第２のＤＡＳＤの記憶ロケーションから、前記第１のデータ・ブロックに対応する第２のデータ・ブロックを読出すステップと、前記第１及び前記第２のデータ・ブロックに対してＸＯＲ演算を実行し、第１のＸＯＲ結果を生成するステップと、前記第１のＸＯＲ結果を第３のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、前記第３のＤＡＳＤをオペレートするステップであって、前記第１のＸＯＲ結果を前記第２のＤＡＳＤから受信するステップと、前記第３のＤＡＳＤの記憶ロケーションから、前記第１及び前記第２のデータ・ブロックに対応する第３のデータ・ブロックを読出すステップと、前記第１のＸＯＲ結果及び前記第３のデータ・ブロックに対してＸＯＲ演算を実行し、第２のＸＯＲ結果を生成するステップと、前記第２のＸＯＲ結果を前記制御装置に転送するステップと、を含む前記オペレーティング・ステップと、を含む、上記（３６）記載の方法。

（３９）前記オペレーティング・ステップが復元オペレーションを含み、該復元オペレーションが、前記第１のＤＡＳＤの前記第１の記憶ロケーションから第１のデータ・ブロックを読出すように、前記第１のＤＡＳＤをオペレートするステップと、第２のＤＡＳＤをオペレートするステップであって、前記第１のデータ・ブロックを前記第１のＤＡＳＤから受信するステップと、前記第２のＤＡＳＤの記憶ロケーションから、前記第１のデータ・ブロックに対応する前記第２のデータ・ブロックを読出すステップと、前記第１及び前記第２のデータ・ブロックに対してＸＯＲ演算を実行し、第１のＸＯＲ結果を生成するステップと、前記第１のＸＯＲ結果を前記第３のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、前記第３のＤＡＳＤをオペレートするステップであって、前記第１のＸＯＲ結果を前記第２のＤＡＳＤから受信するステップと、前記第３のＤＡＳＤの記憶ロケーションから、前記第１及び前記第２のデータ・ブロックに対応する第３のデータ・ブロックを読出すステップと、前記第１のＸＯＲ結果及び前記第３のデータ・ブロックに対してＸＯＲ演算を実行し、第２のＸＯＲ結果を生成するステップと、前記第２のＸＯＲ結果を第４のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、前記第４のＤＡＳＤをオペレートするステップであって、前記第２のＸＯＲ結果を前記第３のＤＡＳＤから受信するステップと、前記第２のＸＯＲ結果を前記第４のＤＡＳＤの記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、を含む、上記（３６）記載の方法。

（４０）前記オペレーティング・ステップが更新オペレーションを含み、該更新オペレーションが、前記第１のデータ・ブロックを前記制御装置から前記第１のＤＡＳＤの前記第１の記憶ロケーションに転送するステップと、前記第１のＤＡＳＤをオペレートするステップであって、前記第１のデータ・ブロックを前記制御装置から受信するステップと、前記第１のＤＡＳＤの前記記憶ロケーションから、前記第１のデータ・ブロックに対応する第２のデータ・ブロックを読出すステップと、前記第１及び前記第２のデータ・ブロックに対してＸＯＲ演算を実行し、第１のＸＯＲ結果を生成するステップと、前記第１のＸＯＲ結果を前記第２のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、前記第２のＤＡＳＤをオペレートするステップであっ

て、前記第１のＸＯＲ結果を前記第１のＤＡＳＤから受信するステップと、前記第２のＤＡＳＤの前記記憶ロケーションから、前記第１及び前記第２のデータ・ブロックに対応する前記第３のデータ・ブロックを読出すステップと、前記第２及び前記第３のデータ・ブロックに対してＸＯＲ演算を実行し、第２のＸＯＲ結果を生成するステップと、前記第２のＸＯＲ結果を前記第３のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、前記第３のＤＡＳＤをオペレートするステップであって、前記第２のＸＯＲ結果を前記第２のＤＡＳＤから受信するステップと、前記第２のＸＯＲ結果を前記第３のＤＡＳＤの記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、を含む、上記（３６）記載の方法。

（４１）前記オペレーティング・ステップがストライプ書込みオペレーションを含み、該ストライプ書込みオペレーションが、第１のデータ・ブロックを前記制御装置から前記第１のＤＡＳＤの前記第１の記憶ロケーションに転送するステップと、前記第１のＤＡＳＤをオペレートするステップであって、前記第１のデータ・ブロックを前記制御装置から受信するステップと、前記第１のデータ・ブロックを前記第１のＤＡＳＤの記憶ロケーションに書込むステップと、前記第１のデータ・ブロックを前記第２のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、前記第２のＤＡＳＤをオペレートするステップであって、前記第２のデータ・ブロックを前記制御装置から受信するステップと、前記第２のデータ・ブロックを前記第２のＤＡＳＤの記憶ロケーションに書込むステップと、前記第１及び前記第２のデータ・ブロックに対してＸＯＲ演算を実行し、第１のＸＯＲ結果を生成するステップと、前記第１のＸＯＲ結果を前記第３のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、前記第３のＤＡＳＤをオペレートするステップであって、前記第３のデータ・ブロックを前記制御装置から受信するステップと、前記第３のデータ・ブロックを前記第３のＤＡＳＤの記憶ロケーションに書込むステップと、前記第１のＸＯＲ結果を前記第２のＤＡＳＤから受信するステップと、前記第１のＸＯＲ結果及び前記第３のデータ・ブロックに対してＸＯＲ演算を実行し、第２のＸＯＲ結果を生成するステップと、前記第２のＸＯＲ結果を前記第４のＤＡＳＤに転送するステップと、を含む前記オペレーティング・ステップと、前記第４のＤＡＳＤをオペレートするステップであって、前記第２のＸＯＲ結果を前記第３のＤＡＳＤから受信するステップと、前記第２のＸＯＲ結果を前記第４のＤＡＳＤの記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、を含む、上記（３６）記載の方法。

【図面の簡単な説明】

【図１】"外部ＸＯＲ"アーキテクチャによるＲＡＩＤ装置１００のブロック図である。

【図２】"順次書込み"オペレーションを実行するために要求されるオペレーションを示す、外部ＸＯＲ　ＲＡＩＤ装置２００のデータ・フロー図である。

【図３】"再生"オペレーションを実行するために要求されるオペレーションを示す、外部ＸＯＲ　ＲＡＩＤ装置２００のデータ・フロー図である。

【図４】本発明によるデータ記憶システムのブロック図である。

【図５】本発明によるＤＡＳＤのハードウェア要素と相互接続を示す詳細ブロック図である。

【図６】本発明による更新タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図７】本発明による再生タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図８】本発明による復元タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図９】本発明によるキャッシュ・データによる更新タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図１０】本発明による故障ＤＡＳＤ状況の更新タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図１１】本発明によるストライプ書込みタスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【符号の説明】

１００　ＲＡＩＤ装置

１０２　中央制御装置

１０４、１０５、１０６、２０４、２０５、２０６、４０４、４０５、４０６、４０７、５００、６０２、６０３、６０４、７０２、７０３、７０４、７０５、８０２、８０３、８０４、８０５、９０２、９０３、９０４、１００２、１００３、１００４、１１０２、１１０３、１１０４　ＤＡＳＤ

１０７、２０８、６０５、９０５、１００５、１１０５　パリティＤＡＳＤ

１１０、４１０　バス

２００　外部ＸＯＲ装置

２０２、４０２、６０１、７０１、８０１、９０１、１００１、１１０１　制御装置

３００、７０６、１００６　故障ＤＡＳＤ

３０８、３０９、３１０、６１０、９１２、９１３、１００９、１１０９、１１１１、１１１３　データ・ブロック

４００、６００、７００、８００、９００、１０００、１１００　データ記憶システム

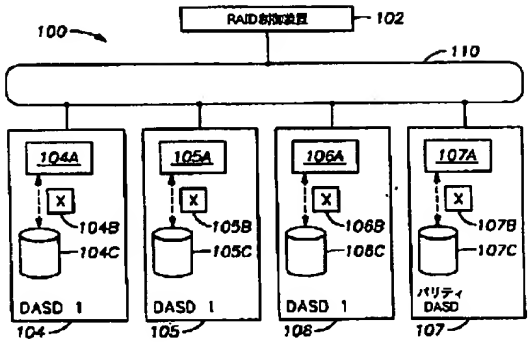
33

411、412、413、414、415、502 インタフェース
418、504 プロセッサ
420、506 データ記憶装置

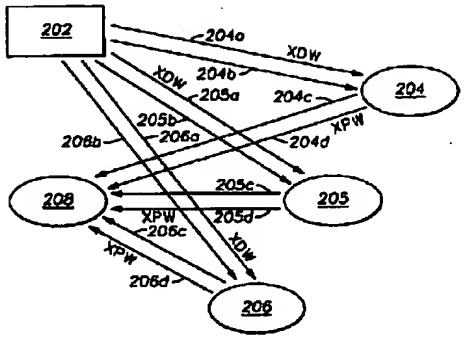
34

422、508、514、516 バッファ
510、512 終端
518 ゲート

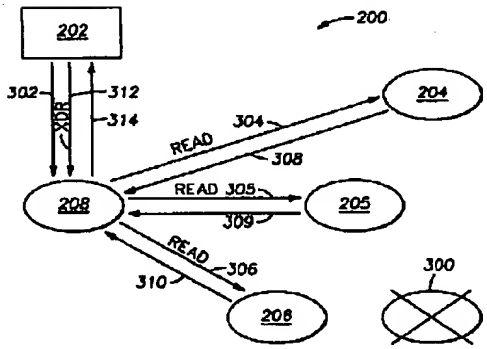
【図１】



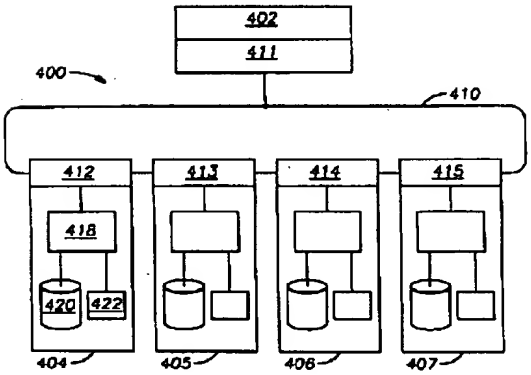
【図２】



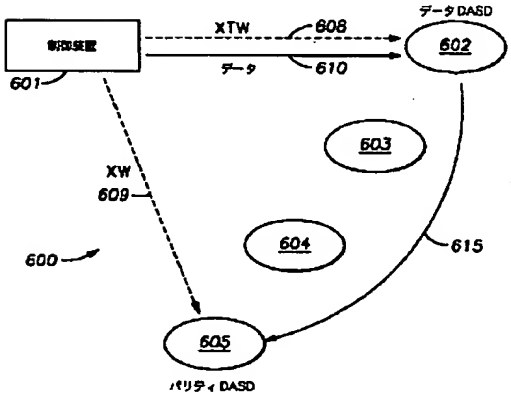
【図３】



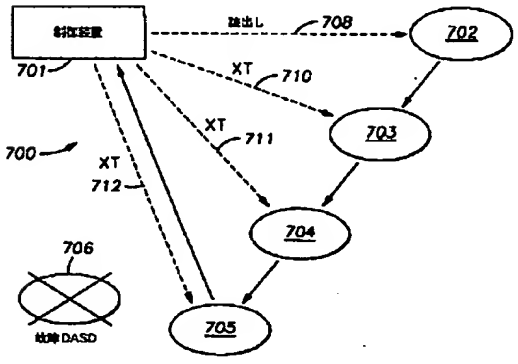
【図４】



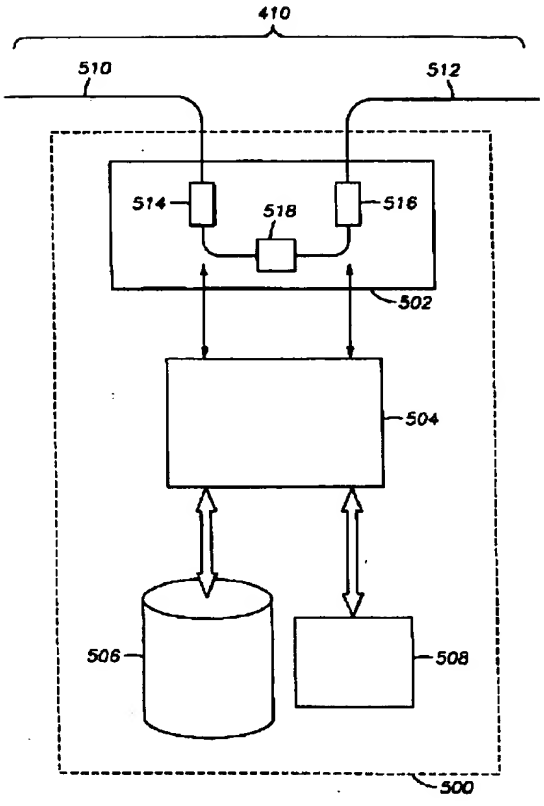
【図６】



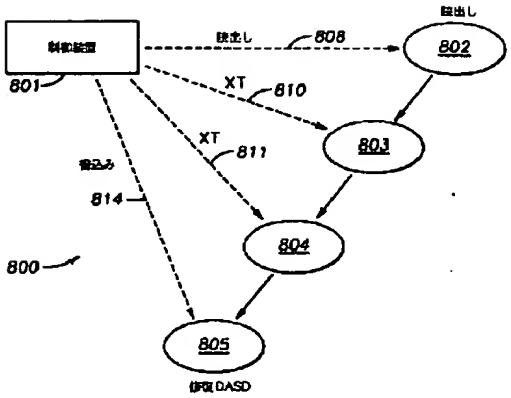
【図７】



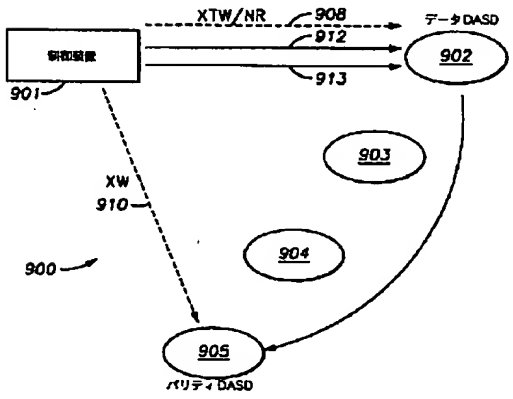
【図５】


410
510
512
514
518
516
502
504
506
508
500

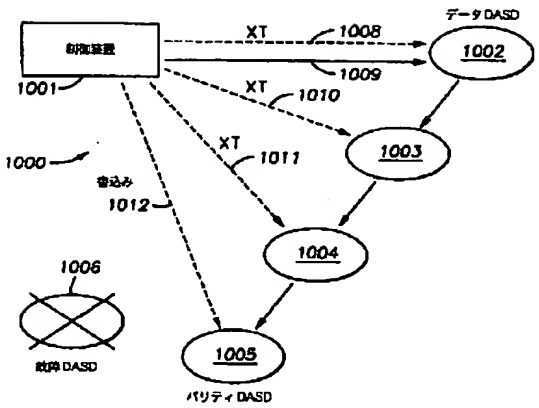
【図８】


801
808
802
XT
810
803
XT
811
814
804
800
805

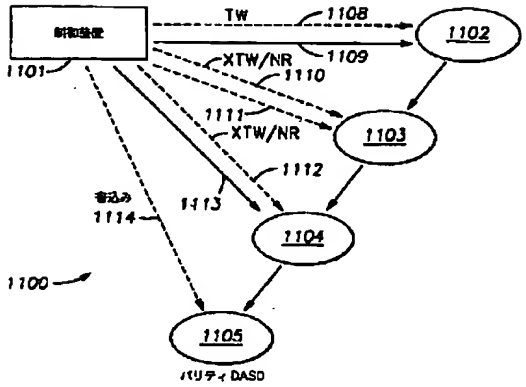
【図９】


XTW/NR
908
912
913
902
901
903
XW
910
904
900
905

【図１０】


XT
1008
1009
1002
1001
XT
1010
1003
XT
1011
1000
1012
1004
1006
1005

【図１１】


制御装置
1101
TW
1108
1102
XTW/NR
1109
1110
1111
XTW/NR
1103
1112
1113
書込み
1114
1104
1100
1105
パリティDASD